

Mode space approach for tight-binding transport simulations in graphene nanoribbon field-effect transistors including phonon scattering

R. Grassi,* A. Gnudi, I. Imperiale, E. Gnani, S. Reggiani, and G. Baccarani
ARCES - DEI, University of Bologna, Viale Risorgimento 2, 40136 Bologna, Italy

(Dated: April 12, 2013)

In this paper, we present a mode space method for atomistic non-equilibrium Green's function simulations of armchair graphene nanoribbon FETs that includes electron-phonon scattering. With reference to both conventional and tunnel FET structures, we show that, in the ideal case of a smooth electrostatic potential, the modes can be decoupled in different groups without any loss of accuracy. Thus, inter-subband scattering due to electron-phonon interactions is properly accounted for, while the overall simulation time considerably improves with respect to real-space, with a speed-up factor of 40 for a 1.5-nm-wide device. Such factor increases with the square of the device width. We also discuss the accuracy of two commonly used approximations of the scattering self-energies: the neglect of the off-diagonal entries in the mode-space expressions and the neglect of the Hermitian part of the retarded self-energy. While the latter is an acceptable approximation in most bias conditions, the former is somewhat inaccurate when the device is in the off-state and optical phonon scattering is essential in determining the current via band-to-band tunneling. Finally, we show that, in the presence of a disordered potential, a coupled mode space approach is necessary, but the results are still accurate compared to the real-space solution.

I. INTRODUCTION

Graphene nanoribbons (GNRs) have been proposed in recent years as a possible replacement for silicon in the future generation of field effect transistors [1, 2]. Despite the outstanding challenges in producing GNRs with controlled width and edges and the resulting degradation of the graphene intrinsic mobility, devices with large on-off current ratio have been successfully demonstrated [3].

Theoretical performance of GNR-FETs have been widely investigated by numerous simulation studies (e.g Refs. 4–9). The state-of-the-art approach for modeling the electronic properties of GNRs is based on an atomistic tight-binding (TB) Hamiltonian with a p_z orbital basis set. The transport problem is usually solved within the nonequilibrium Green's function formalism (NEGF) [10], which provides a rigorous framework for including incoherent scattering processes in the quantum description. TB simulations of GNR-FETs including phonon scattering have been reported too [11–14]. Those simulations were performed using a real space (RS) approach. On the other hand, more efficient mode space (MS) methods would be preferable for use in intensive device simulations.

The MS approach is well established with reference to an effective mass (EM) Hamiltonian [15, 16]. It is based on the expansion of the Green's functions in terms of the transverse eigenfunctions (modes) and it is most efficient when quantum confinement is relatively strong in the transverse plane, so that only few of the lowest subbands are occupied. The so-called coupled mode space (CMS) is the general method, while a more efficient version, the uncoupled mode space (UMS), can be adopted

when the transverse potential profile is slightly varying along the transport direction. The inclusion of electron-phonon scattering in the MS EM model is also well known [17–21].

On the other hand, the MS approach is not generally applicable to a TB Hamiltonian. Due to the non-separable Hamiltonian, even when the potential is uniform along the longitudinal direction (so that the longitudinal wavevector k is conserved), the wavefunctions at different k in the same subband are in general different. In other words, by defining the modes as the transverse part of the wavefunctions at a particular k , the wavefunctions at the generic k are a linear combination of them. This is similar to what occurs in the case of a $k \cdot p$ Hamiltonian [22]. Nevertheless, for the particular case of the TB Hamiltonian of GNRs with armchair edges, we have previously shown by numerical calculation that a MS method is actually possible, since each subband contains only few modes [23]. A formal derivation was given in Ref. 24 using analytically defined modes.

In this paper, we propose a novel MS method for the inclusion of graphene acoustic phonon (AP) and optical phonon (OP) scattering within the NEGF formalism, and test its accuracy with respect to RS in both the cases of smooth and disordered potentials. Two approximations commonly found in the literature for treating the scattering self-energy are also evaluated. The paper is organized as follows. Sec. II reviews the standard RS TB method as well as the MS TB method from Ref. 23, modified so as to include phonon scattering. Simulation results of GNR-FETs are presented in Sec. III and conclusions are finally drawn in Sec. IV.

* rgrassi@arces.unibo.it

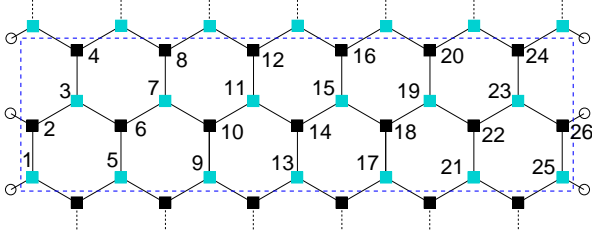


FIG. 1. One-dimensional elementary cell of an $N_a = 13$ armchair GNR.

II. MATHEMATICAL MODEL

A. Real space formulation

The atomic structure of an armchair GNR unit cell is shown for reference in Fig. 1. The GNR width is denoted by the number of dimer lines N_a . We adopt the TB Hamiltonian model proposed in Ref. 25, which is based on a set of orthogonal p_z orbitals, one for each atom, with hopping integrals limited to first nearest-neighbor orbitals and of value t for internal atom pairs, and $t(1+\delta)$ for atom pairs located along the edges of the GNR ($t = -2.7$ eV and $\delta = 0.12$). The onsite energies are set equal to the value of the electrostatic potential energy at each atomic site. Unless stated otherwise, the electrostatic potential is calculated by self-consistently solving the 3D Poisson equation.

In order to calculate electron and hole densities within the NEGF formalism, the following equations need to be solved for the retarded, the lesser and the greater Green's function matrices \mathbf{G}^R , $\mathbf{G}^<$ and $\mathbf{G}^>$, respectively, for each energy E

$$[(E + i0^+)\mathbf{I} - \mathbf{H}^d - \boldsymbol{\Sigma}^R(E)] \mathbf{G}^R(E) = \mathbf{I} \quad (1)$$

$$\mathbf{G}^<(E) = \mathbf{G}^R(E) \boldsymbol{\Sigma}^<(E) \mathbf{G}^A(E) \quad (2)$$

$$\mathbf{G}^>(E) = \mathbf{G}^R(E) \boldsymbol{\Sigma}^>(E) \mathbf{G}^A(E) \quad (3)$$

where \mathbf{H}^d denotes the restriction of the Hamiltonian matrix \mathbf{H} to the GNR portion inside the simulation domain and $\mathbf{G}^A = \mathbf{G}^{R\dagger}$. Besides, $\boldsymbol{\Sigma}^R(E) = \boldsymbol{\Sigma}_P^R(E) + \boldsymbol{\Sigma}_S^R(E) + \boldsymbol{\Sigma}_D^R(E)$ and similarly for $\boldsymbol{\Sigma}^>(E)$ and $\boldsymbol{\Sigma}^<(E)$, where $\boldsymbol{\Sigma}_{S/D}^{(\dots)}$ are the self-energies for the semi-infinite source and drain leads, which are assumed in thermodynamic equilibrium, while $\boldsymbol{\Sigma}_P^{(\dots)}$ account for phonon scattering. Calculation of both (2) and (3) is actually not needed, since either one of the two equations can be replaced by the identity $\mathbf{G}^> - \mathbf{G}^< = \mathbf{G}^R - \mathbf{G}^A$.

Phonon scattering is treated within the self-consistent Born approximation. The model accounts for both OP and AP scattering, the latter within the elastic and high temperature limit, resulting in the following expressions

of the lesser and greater phonon self energies

$$\boldsymbol{\Sigma}_P^<(E) = [D_{ap} \mathbf{G}^<(E) + D_{op}(N_{op} + 1) \mathbf{G}^<(E + \hbar\omega_{op}) + D_{op} N_{op} \mathbf{G}^<(E - \hbar\omega_{op})] \circ \mathbf{I} \quad (4)$$

$$\boldsymbol{\Sigma}_P^>(E) = [D_{ap} \mathbf{G}^>(E) + D_{op}(N_{op} + 1) \mathbf{G}^>(E - \hbar\omega_{op}) + D_{op} N_{op} \mathbf{G}^>(E + \hbar\omega_{op})] \circ \mathbf{I} \quad (5)$$

where \circ indicates element-by-element matrix multiplication, N_{op} is the OP occupation number, and D_{ap} and D_{op} are given by

$$D_{ap} = \frac{D_{ac}^2 k_B T}{4m_c v_s^2}, \quad D_{op} = \frac{D_o^2 \hbar}{8m_c \omega_{op}} \quad (6)$$

with $D_{ac} = 16$ eV the AP deformation potential, $v_s = 2 \times 10^6$ cm/s the sound velocity in graphene, T the temperature, m_c the carbon atomic mass, $D_o = 10^9$ eV/cm and $\hbar\omega_{op} = 160$ meV the zone-boundary OP deformation potential and energy, respectively. The retarded phonon self-energy is computed using the identity

$$\begin{aligned} \boldsymbol{\Sigma}_P^R(E) &= \frac{i}{2\pi} \int_{-\infty}^{+\infty} \frac{\boldsymbol{\Sigma}_P^>(E') - \boldsymbol{\Sigma}_P^<(E')}{E + i0^+ - E'} dE' = \\ &= \frac{\boldsymbol{\Sigma}_P^>(E) - \boldsymbol{\Sigma}_P^<(E)}{2} + i\mathcal{P} \int_{-\infty}^{+\infty} \frac{\boldsymbol{\Sigma}_P^>(E') - \boldsymbol{\Sigma}_P^<(E')}{2\pi(E - E')} dE' \end{aligned} \quad (7)$$

where the symbol \mathcal{P} stands for the principal part of the integral. Unless stated otherwise, we include the contribution to the principal part integral (Hermitian part of $\boldsymbol{\Sigma}_P^R$) due to AP scattering, while we neglect the one due to OP scattering. See Appendix A for details on the numerical implementation of the principal part integral.

The equations above, namely (1)-(3) together with (4), (5) and (7), are iteratively solved. It is worth noticing that \mathbf{H}^d and $\boldsymbol{\Sigma}^{(\dots)}$ can be ordered in a block tridiagonal and block diagonal form, respectively, each diagonal block having the size of the number of atoms in a specific transverse layer, or slab, of the GNR. Hence, the Recursive Green Function algorithm (RGF) [26, 27] can be used to solve (1)-(3). In this paper, we consider each slab to be made of four rows of atoms, i.e. equal to the GNR unit cell (Fig. 1).

B. Mode space formulation

In this section the MS approach presented in Ref. 23, and in that paper discussed with reference to coherent transport simulations of GNRs, is extended to include phonon scattering.

The MS formulation starts by defining a set of orthonormal vectors, or modes, $\phi(i)$ for each slab i of the device. We choose the modes of the generic slab as the eigenvectors computed at $k = 0$ of a fictitious GNR obtained by the periodic repetition (potential energy included) of that slab along the longitudinal direction. In

the absence of topological differences between the slabs (such as edge irregularities or internal vacancies), the modes are the solution of the eigenvalue problem

$$\left(\mathbf{H}_{i,i} + \mathbf{H}_{i,i+1} + \mathbf{H}_{i,i+1}^\dagger\right) \phi^m(i) = \varepsilon^m \phi^m(i) \quad (8)$$

where $\mathbf{H}_{i,j}$ is the Hamiltonian block between slab i and j . Denoting by $\mathbf{v}(i)$ the matrix whose columns are the modes of layer i , i.e. $\mathbf{v}(i) = [\phi^1(i) \cdots \phi^m(i) \cdots]$, a block-diagonal transformation matrix \mathbf{V} satisfying $\mathbf{V}^\dagger \mathbf{V} = \mathbf{I}$ is constructed

$$\mathbf{V} = \begin{pmatrix} \mathbf{v}(1) & 0 & \cdots & 0 \\ 0 & \mathbf{v}(2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \cdots & \ddots & \mathbf{v}(N_s) \end{pmatrix} \quad (9)$$

with N_s equal to the number of slabs.

In the following, we indicate with a tilde the MS matrices to distinguish them from the RS matrices. The CMS method consists in approximating \mathbf{G}^R (and similarly for $\mathbf{G}^<$ and $\mathbf{G}^>$) as

$$\mathbf{G}^R(E) \simeq \mathbf{V} \tilde{\mathbf{G}}^R(E) \mathbf{V}^\dagger \quad (10)$$

where $\tilde{\mathbf{G}}^R$ is the solution of

$$\left[(E + i0^+) \mathbf{I} - \tilde{\mathbf{H}}^d - \tilde{\Sigma}^R(E)\right] \tilde{\mathbf{G}}^R(E) = \mathbf{I} \quad (11)$$

with

$$\tilde{\mathbf{H}}^d = \mathbf{V}^\dagger \mathbf{H}^d \mathbf{V} \quad (12)$$

$$\tilde{\Sigma}^R(E) = \mathbf{V}^\dagger \Sigma^R(E) \mathbf{V} \quad (13)$$

Eq. (11) is the MS version of (1). Analogous considerations apply to the other equations. Eq. (10) would be exact if \mathbf{V} was a square matrix. In practice, a mode truncation is performed so that $\tilde{\mathbf{H}}^d$ and $\tilde{\Sigma}^R$ have a smaller size than their corresponding RS matrices and the solution of (11) instead of (1) is computationally advantageous. The computational time of the RGF algorithm scales as $O(N_y^3 N_s)$, where N_y is the matrix block size [27], which is equal to the slab size for RS ($2N_a$ according to our choice of the slab size) and to the number N_m of selected modes per slab for CMS. Thus, the speed-up of CMS compared to RS is a factor of the order of $(2N_a/N_m)^3$.

An algorithm to select the modes to retain in the \mathbf{V} matrix was presented in Ref. 23. It is based on identifying, among the modes calculated with zero electrostatic potential, the minimum set of modes that allow to reproduce with sufficient accuracy the subbands that lie in the energy range of interest. The mode indexes so identified are then used to select the actual modes calculated with the non-null electrostatic potential. As shown in Fig. 2 for a $N_a = 13$ GNR, it turns out that each one of the lowest conduction subbands at zero potential can be well

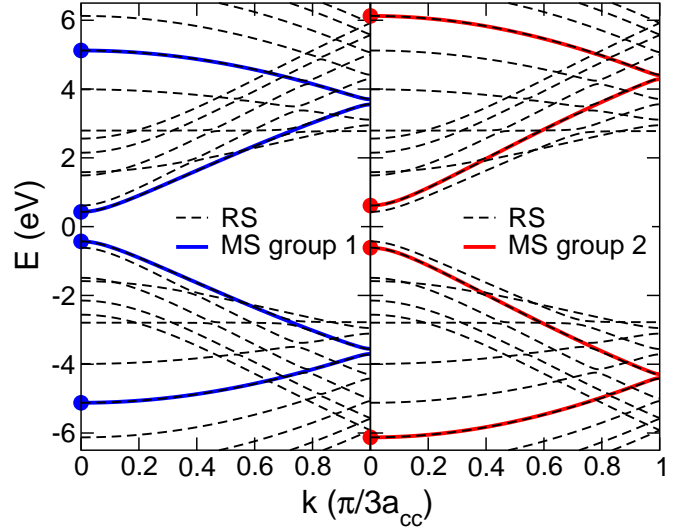


FIG. 2. Subband structure of a $N_a = 13$ GNR computed with RS (dashed lines) and MS (solid lines) using two different groups of modes (left and right). The modes included in each group are the ones that correspond to the eigenvalues at $k = 0$ indicated with circles in each figure. The electrostatic potential is set to zero. $3a_{cc}$ is the length of the GNR unit cell, with a_{cc} the carbon interatomic distance.

reproduced by just four modes, which correspond to the eigenvalues at $k = 0$ belonging to (i) the considered subband, (ii) the valence subband symmetrical to it, and (iii-iv) their respective folded continuations. Symmetrical considerations apply to the highest valence subbands. GNRs with different N_a behave similarly. Since the sets of four modes described above, hereafter referred to as groups of modes, are disjoint from each other, a more efficient MS method, here called uncoupled group mode space (UGMS), can be obtained from (10)-(13) by neglecting the coupling between modes belonging to different groups. Let $\mathbf{v}_b(i)$ be the matrix whose columns are the modes in layer i of group b only. A transformation matrix \mathbf{V}_b similar to (9) is constructed by using the matrices \mathbf{v}_b as diagonal blocks. The UGMS method is derived by further approximating (10) as

$$\mathbf{G}^R(E) \simeq \sum_{b=1}^{N_g} \mathbf{V}_b \tilde{\mathbf{G}}_b^R(E) \mathbf{V}_b^\dagger \quad (14)$$

where N_g is the number of the considered groups and $\tilde{\mathbf{G}}_b^R$ is the solution of

$$\left[(E + i0^+) \mathbf{I} - \tilde{\mathbf{H}}_b^d - \tilde{\Sigma}_b^R(E)\right] \tilde{\mathbf{G}}_b^R(E) = \mathbf{I} \quad (15)$$

with

$$\tilde{\mathbf{H}}_b^d = \mathbf{V}_b^\dagger \mathbf{H}^d \mathbf{V}_b \quad (16)$$

$$\tilde{\Sigma}_b^R(E) = \mathbf{V}_b^\dagger \Sigma^R(E) \mathbf{V}_b \quad (17)$$

In the UGMS method, the computational cost of the RGF algorithm scales as $O(N_g N_y^3 N_s)$, where $N_y = 4$

is the number of modes in each group. Hence, if the same number of modes $N_m = 4N_g$ is used, the speed-up of UGMS compared to CMS scales as N_g^2 , which can be sizeable for wide ribbons, since N_g is proportional to the number of subbands that contribute to transport, which, in turn, is proportional to the GNR width.

In the following, we refer to the UGMS formulation, since the CMS one can be recovered from Eqs. (14)-(17) by considering all the selected modes as belonging to the same group and by setting $N_g = 1$. Note that Σ_S^R and Σ_D^R can be directly computed in MS without making use of (17). Instead, the calculation of $\tilde{\Sigma}_P^{(\dots)}$ is more complicated. In particular, in the MS representation these matrices are no longer diagonal, but just block diagonal. For example, as far as $\tilde{\Sigma}_P^<$ is concerned, from the equation for $\tilde{\Sigma}^<$ analogous to (17), using (4) and considering for simplicity the contribution of APs only (first term at the right-hand-side), with $\mathbf{G}^<$ replaced by the equation analogous to (14), the following MS expression for the element relative to modes m and m' of the diagonal matrix block associated with slab i and group b is derived

$$\tilde{\Sigma}_{b,mm'}(i, i; E) = D_{ap} \sum_{b', n, n'} F_{b,mm'}^{b', nn'}(i) \tilde{G}_{b', nn'}(i, i; E) \quad (18)$$

where the symbols $<$ and P have been dropped for brevity and the form factor F is defined by

$$F_{b,mm'}^{b', nn'}(i) = \sum_{\alpha} v_{b, \alpha m}^*(i) v_{b, \alpha m'}(i) v_{b', \alpha n}(i) v_{b', \alpha n'}^*(i) \quad (19)$$

(n and n' are mode indexes within group b' , α the atom index within the slab). Identical considerations apply to the OP terms and to the other types of phonon self-energies. It should be noticed that *all the considered modes* give their contribution to $\tilde{\Sigma}_{b,mm'}$, not only those belonging to the same group b of modes m, m' . This means that intersubband scattering between all considered modes is taken into account, without the necessity of going to full RS simulations. It must also be noticed that in the solution process the different mode groups can be treated independently, which is a great advantage compared to RS and also to CMS: at every iteration pass, once the Green functions for all groups have been calculated, the self-energies are updated using the equations like (18) collecting the contributions from all modes.

In the next section the MS and RS solutions will be compared in the presence of phonon scattering. A further approximation will be tested, which consists in simplifying the form factor as

$$F_{b,mm'}^{b', nn'}(i) \simeq \delta_{m, m'} \delta_{n, n'} \sum_{\alpha} |v_{b, \alpha m}(i)|^2 |v_{b', \alpha n}(i)|^2 \quad (20)$$

where δ is the Kronecker delta, so that only the diagonal entries are kept in (18) and similar equations. This approximation is usually adopted in the context of the CMS EM approach due to its higher efficiency [17, 18, 20]. The results obtained with expressions (19) and (20) will be named “MS2” and “MS1”, respectively.

III. RESULTS

All simulated devices have a double-gate structure with SiO₂ 1-nm-thick top and bottom oxides. As mentioned before, the source and drain are assumed to be semi-infinite leads. The portions of the source and drain regions included in the simulation domain are 10 nm long and uniformly doped. The channel is assumed to be intrinsic. Unless noted otherwise, the following device parameters are used: $N_a = 13$ (corresponding to a ribbon width $W \simeq 1.5$ nm), gate length $L_G = 17$ nm, and doping concentration in the source and drain regions equal to 10^{-2} dopants per carbon atom. Also, unless stated otherwise, the MS method simulations are performed using the uncoupled group approximation (UGMS). For the reference $N_a = 13$ GNR, the 2 groups of 4 modes of Fig. 2 are used (8 modes out of a total of 26 that correspond to the full RS solution). Both conventional FET and tunnel FET (TFET) devices are simulated. The latter have received great attention in recent years for their potential in low-power applications [8]. The two structures only differ in the type of doping of the source region. While the drain is always n-type, the source is n-type and p-type for conventional FETs and TFETs, respectively.

We consider only ideal GNRs with perfect edges and no internal vacancies. Nevertheless, in the following, we also study the effect of adding a disorder potential to the TB Hamiltonian of the ideal GNR. Such disorder potential can, in principle, mimic the perturbation effect of impurities or the substrate.

A. Smooth electrostatic potential

The case of no disorder potential is treated first.

The I vs. V_{GS} (“turn-on”) characteristics of a conventional n-i-n FET at $V_{DS} = 0.1$ and 0.8 V are plotted in Fig. 3. They are computed with the RS and the MS methods. For each method, we consider three transport conditions: (i) without scattering (i.e. ballistic transport), (ii) in the presence of only AP scattering, and (iii) in the presence of both AP and OP scattering. In the case of phonon scattering and MS approach, the simplified expression of the form-factor in (20) is used (MS1).

As regards the effect of phonon scattering and with reference to the RS results, it can be seen that AP scattering has only a limited effect on the current at this channel length, resulting in a ballisticsity ratio (i.e. ratio between current in the presence of phonon scattering and ballistic current) of 0.8 at $V_{GS} = 0.8$ V for both V_{DS} values. When also OP scattering is included, the current at high V_{GS} , i.e. on-state current, is only slightly decreased for the largest V_{DS} value. Similar findings were reported in Ref. 12. On the other hand, OP scattering is responsible for an increase of the minimum off-state current by a few orders of magnitudes when V_{DS} is low. This effect is caused by energy relaxation through emission or absorption of optical phonons, which favors band-to-band

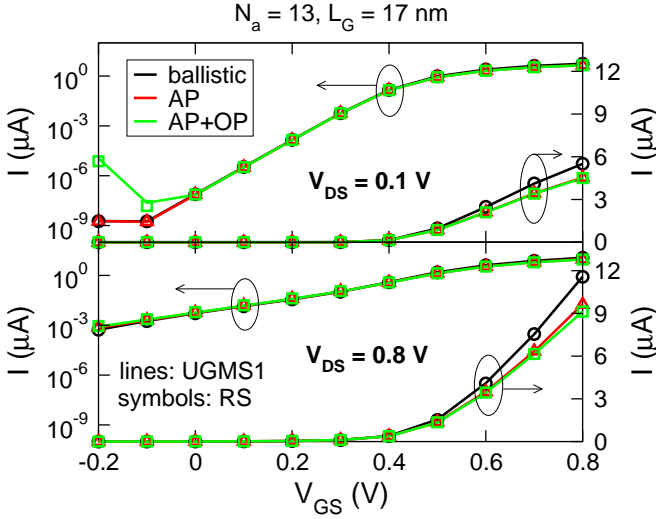


FIG. 3. Turn-on characteristics of a n-i-n FET with $W = 1.5$ nm at $V_{DS} = 0.1$ V (top) and $V_{DS} = 0.8$ V (bottom). Results are obtained by the RS and UGMS1 methods for different transport conditions: ballistic transport, AP scattering, and both AP and OP scattering.

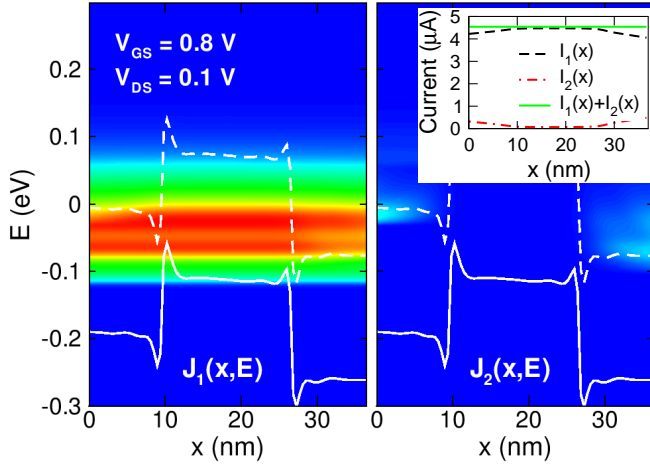


FIG. 4. Current spectrum $J_b(x, E)$ obtained by UGMS1 method for mode group $b = 1$ (left) and $b = 2$ (right) for the device in Fig. 3 with AP scattering and at the bias indicated in the figure. Blue (red) color means low (high) density. The white solid (dashed) line is the profile of the first (second) conduction subband. The source Fermi level is at $E = 0$. The inset shows the integrals over energy $I_b(x) = \int J_b(x, E) dE$ and the conservation of the total current $I_1(x) + I_2(x)$.

tunneling (BTBT) and shifts toward positive V_{GS} values the onset of the ambipolar conduction by BTBT, similar to what occurs in carbon nanotubes [28].

With regard to the MS results, we do not observe any significant discrepancy with respect to RS, except for the curve with OP scattering at $V_{DS} = 0.1$ V, close to the V_{GS} point of minimum current. This discrepancy will be discussed later in the text.

In order to test the validity of the MS method in de-

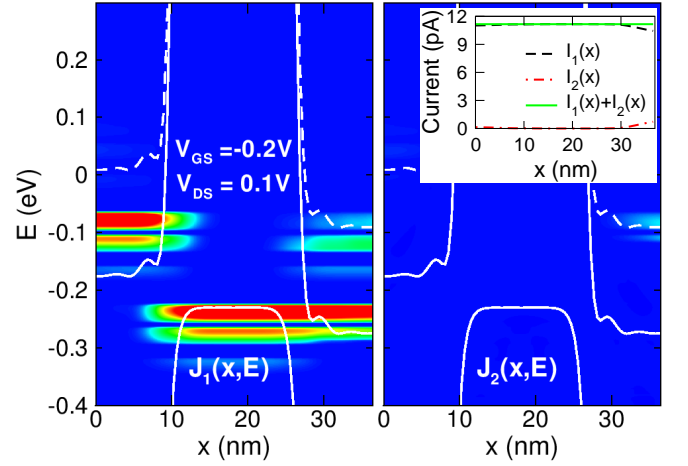


FIG. 5. Same as in Fig. 4 but with both AP and OP scattering and at $V_{GS} = -0.2$ V and $V_{DS} = 0.1$ V. The white solid (dashed) lines are the profiles of the first (second) pairs of conduction and valence subbands.

scribing inter-subband scattering, we separately plot in Fig. 4 the MS current spectrum for the first and second group of modes (corresponding to the first and second subband, respectively) at $V_{GS} = 0.8$ V and $V_{DS} = 0.1$ V, in the case with only AP scattering. It can be seen that both subbands contribute to current. In addition, despite the absence of inelastic scattering processes in the simulation, the distribution of current over energy of the single subband is not conserved when moving from source to drain, indicating that some of the carriers are transferred from one subband to the other. Indeed, thanks to phonon scattering, part of the electrons injected from the source in the second subband reach the drain by traversing the channel in the first subband, where the energy barrier is lower. The total current is properly conserved as shown in the inset of Fig. 4. Inter-subband scattering and the conservation of the total current are also evident from simulations including OP scattering: see Fig. 5, which corresponds to a bias point where transport is dominated by phonon-assisted BTBT.

Next, we consider the TFET architecture. Fig. 6-top shows the I vs. V_{GS} characteristics at $V_{DS} = 0.4$ V for the p-i-n counterpart of the device in Fig. 3, calculated with the different methods and by including different types of scattering, as indicated in the legend. Results obtained with both expressions MS1 and MS2 of the form-factor are reported. The symmetry of the characteristics is related to the symmetric doping of the source and drain regions [8]. By looking at the RS results, one can see that AP scattering has a negligible effect in both the on- and off-state regimes, while OP scattering significantly increases the minimum leakage current and, consequently, the minimum inverse subthreshold slope (SS), similar to the n-i-n case. As illustrated by the current spectrum in Fig. 7-left, the increase of the minimum current is due to phonon-assisted BTBT at the source-

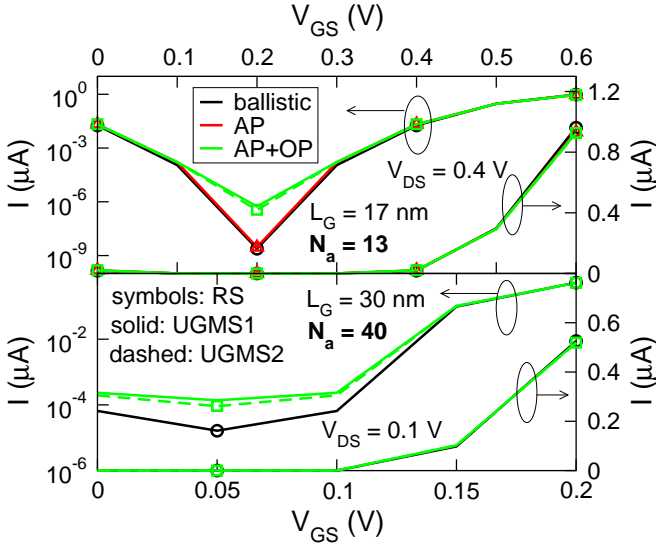


FIG. 6. Turn-on characteristics of p-i-n FETs. Top: device with $W = 1.5$ nm at $V_{DS} = 0.4$ V. Bottom: device with $W = 5$ nm at $V_{DS} = 0.1$ V. Results are obtained by the RS and UGMS methods for different transport conditions: ballistic transport, AP scattering (only for the device with $W = 1.5$ nm), and both AP and OP scattering. In the MS case, we compare the results obtained by expressions MS1 and MS2 of the form-factor.

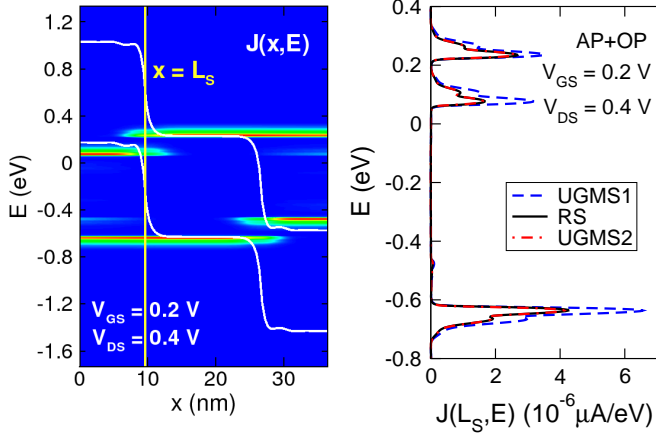


FIG. 7. Left: current spectrum $J(x, E)$ obtained by RS method for the device in Fig. 6-top with both AP and OP scattering and at the bias indicated in the figure. The vertical line indicates the position L_S of the source-channel junction. The two white lines are the profile of the first conduction and valence subbands. Right: comparison between the current spectra at $x = L_S$ obtained by RS, UGMS1, and UGMS2 methods.

channel and drain-channel junctions: although the first conduction and valence subbands do not face each other, electrons can transmit from the valence to the conduction subband by absorption of optical phonons. These results are in agreement with the ones in Ref. 13.

Again, the only difference (of about a factor of 1.5) between the RS and MS1 results is noticed in the pres-

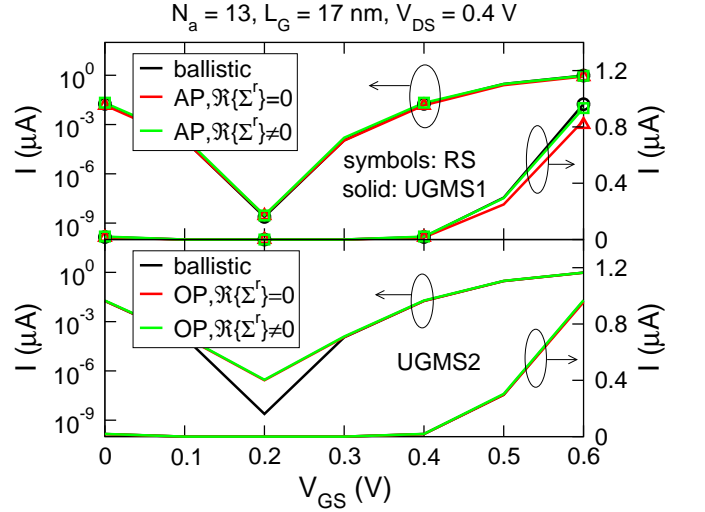


FIG. 8. Turn-on characteristics for the same device as in Fig. 6-top comparing different approximations for Σ_P^R . Top: only AP scattering is included, with or without the Hermitian part of Σ_P^R . Bottom: only OP scattering is included, with or without the Hermitian part of Σ_P^R . The ballistic curve is shown for reference in both figures and the solution method for each curve is indicated in the legend.

ence of OP scattering at the point of minimum current. However, the accuracy with respect to RS can be almost completely recovered by the MS2 method, which uses the exact expression of the form-factor in (19). This is more clearly shown by the comparison in Fig. 7-right between the current spectra at the source-channel junction obtained with the different methods. The lack of accuracy of (20) in this bias condition could be related to the neglect of the terms $F_{b,mn}^{b,mn}(i)$ with $n \neq m$, which are actually of the same size as the terms $F_{b,mm}^{b,nn}(i)$ included in (20). As a drawback, a slow-down of the simulation by about a factor of 1.5 has been measured using MS2 compared to MS1, which can be ascribed to the increased computational cost of (18).

The above considerations apply also to the turn-on characteristics of a wider GNR TFET in Fig. 6-bottom. For this device, a lower V_{DS} is chosen due to the lower band gap. In addition, we set $N_a = 40$ (corresponding to $W \simeq 5$ nm), $L_G = 30$ nm, and a doping concentration in the source and drain regions of $7 \cdot 10^{-4}$ dopants/atom. 4 groups of 4 modes are used in the UGMS simulations. It is worth noticing that the UGMS method is still accurate for the wider ribbon, indicating that the decoupling in separate groups is still valid, even though the subbands are more closely spaced than in the $N_a = 13$ GNR.

To evaluate the importance of the Hermitian part of Σ_P^R , here denoted by $\Re\{\Sigma_P^R\}$, we report in Fig. 8 the I vs. V_{GS} characteristics of the $N_a = 13$ TFET computed with or without $\Re\{\Sigma_P^R\}$, separately for each type of scattering.

For AP scattering (Fig. 8-top), the neglect of $\Re\{\Sigma_P^R\}$ leads to an underestimation of the on-state current. This can be understood as follows. In general, $\Re\{\Sigma_P^R\}$ has the

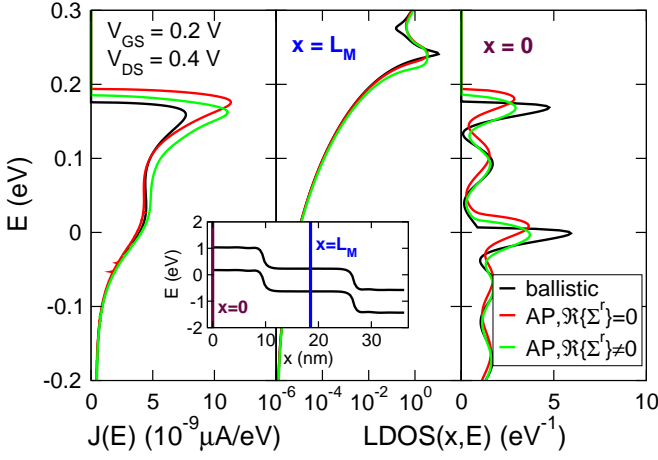


FIG. 9. Current spectrum (left) and integral over the slab of the LDOS at $x = L_M$ (center) and $x = 0$ (right) for the device in Fig. 8–top at $V_{GS} = 0.2$ V and $V_{DS} = 0.4$ V. L_M is the mid-channel position. RS method is used. The ballistic subband profile is shown in the inset.

effect of shifting the Hamiltonian eigenvalues [29]. However, in GNRs, the shift is of opposite sign for energies above and below the GNR mid-gap due to the symmetry of the subband structure. The result is a decrease of the GNR band gap, which favors BTBT, so that a larger current is expected when $\Re\{\Sigma_P^R\}$ is included in the simulation. Interestingly, the two expressions of $\Re\{\Sigma_P^R\}$ give almost identical values with respect to the minimum current. To clarify this point, we plot in Fig. 9 the current spectrum and the local density of states (LDOS) per slab at two positions along the device, for the bias point corresponding to the minimum current. First, it can be seen that the current spectrum with AP scattering and $\Re\{\Sigma_P^R\} \neq 0$ is larger than the ballistic one: the reason can be ascribed to an enhanced BTBT through the channel region due to the band gap narrowing effect mentioned above, which can be appreciated from the logarithmic plot of the LDOS at the mid-channel position in Fig. 9–center. Secondly, by looking at Fig. 9–right, it can be noticed that the peaks of the LDOS in the source region with $\Re\{\Sigma_P^R\} \neq 0$ are located at the same energy positions as the ones of the ballistic LDOS. On the contrary, the peaks of the LDOS with $\Re\{\Sigma_P^R\} = 0$ are shifted up in energy, resulting in a tunneling current larger than the ballistic one and similar to the one with $\Re\{\Sigma_P^R\} \neq 0$ (Fig. 9–left). The shift of the LDOS can be attributed to a “loss of charge” when $\Re\{\Sigma_P^R\}$ is set to zero [30] and to the combined effect of the electrostatic feedback.

For OP scattering instead, no relevant difference is observed when including $\Re\{\Sigma_P^R\}$, even at high V_{GS} (Fig. 8–bottom).

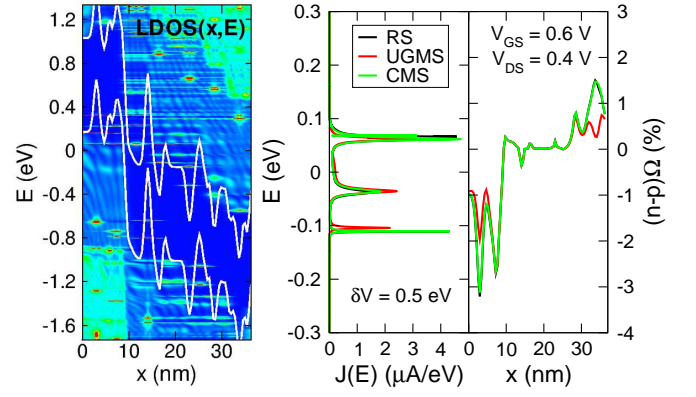


FIG. 10. Integral over the slab of the LDOS (left), current spectrum (center), and average over the slab of the net electron concentration per carbon atom (right) for the device in Fig. 6–top at $V_{GS} = 0.6$ V and $V_{DS} = 0.4$ V, in the presence of a long-range disorder potential and in the coherent limit. LDOS is obtained by RS, whereas the solution method for the other curves is indicated in the legend. $\Omega = 3a_{cc}^2/4$ is half the area of the graphene unit cell.

B. Disordered electrostatic potential

We focus on the $N_a = 13$ TFET. The simulations are performed in a non-self-consistent way, by solving the NEGF equations with a fixed electrostatic potential. We take the electrostatic potential as the sum of a disorder potential and the one calculated self-consistently in the absence of disorder and in the ballistic limit at $V_{GS} = 0.6$ V and $V_{DS} = 0.4$ V, using the RS approach. Two types of disorder are considered: a “long-range” one, i.e. slowly varying on the atomic scale, and a “short-range” one, i.e. rapidly varying from one atom to its neighbor ones. According to the first model, which is derived from Ref. 31, the disorder energy potential V_i at the atomic site i , located at position \vec{r}_i , is calculated as

$$V_i = \sum_{j=1}^N S_j \delta V \exp\left(-\frac{|\vec{r}_i - \vec{X}_j|^2}{l^2}\right) \quad (21)$$

where N , l , δV are parameters, while S_j and \vec{X}_j are random variables: S_j can take values ± 1 with equal probability; \vec{X}_j is uniformly distributed over all the atomic position \vec{r}_i . The second model we study is the Anderson type of disorder [32], according to which $V_i = Y_i$, where Y_i is a random variable uniformly distributed in $[-\delta V/2, \delta V/2]$.

The coherent case, i.e. without phonon scattering, is considered first. In Fig. 10–left we show the LDOS corresponding to a realization of the long-range type of disorder, obtained with $N = 0.01N_c$, $l = 5a_{cc}$, and $\delta V = 0.5$ eV, where N_c is the total number of carbon atoms inside the device and a_{cc} is the carbon-carbon bond length. The resonant states induced by disorder are clearly visible in the LDOS. In Fig. 10–center and

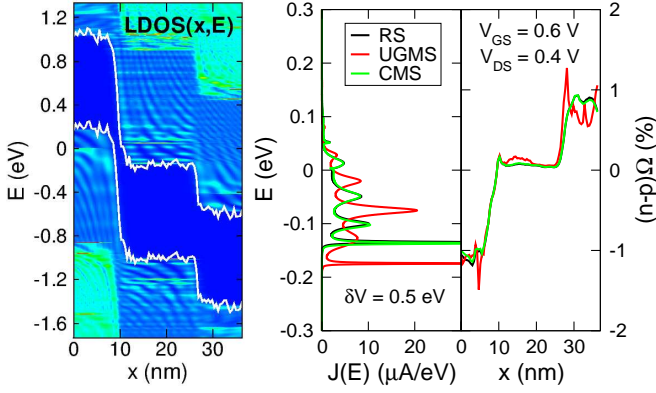


FIG. 11. Same as in Fig. 10 but in the presence of a short-range disorder potential.

Fig. 10–right we compare the current spectrum and net electron charge along the device, respectively, obtained with RS, UGMS, and CMS. The UGMS simulations are performed with the same 2 groups of 4 modes used in the case without disorder, while the CMS ones use the same 8 modes but all coupled in one group. It turns out that the UGMS method loses accuracy due to disorder-induced mode mixing, while the CMS method leads to results very close to the RS ones (e.g. the error on the current value is less than 2%).

Similar considerations can be made regarding the results obtained with the second model of disorder with $\delta V = 0.5$ eV. (Fig. 11). The more regular LDOS pattern and the higher current spectrum in Fig. 11 compared to Fig. 10 indicate that the amount of scattering in this device is smaller.

The simulations have been repeated including AP and OP scattering (Fig. 12). For both types of disorder, it is seen that the MS1 and MS2 approximations of the form-factor provide similar results and that the CMS approach is still accurate with respect to RS. By comparing the current spectra in Fig. 12 with the ones in Figs. 10–11, it can be noticed that phonon scattering adds a significant broadening. In the case of short-range disorder, phonon scattering only slightly decreases the current (from 0.83 to 0.80 μA). On the other hand, in the case of long-range disorder, the current is increased from 0.084 to 0.10 μA when phonon scattering is included in the simulation, indicating that the localization transport regime [29] that occurs in the coherent approximation is broken by the dephasing effect of phonon scattering.

IV. CONCLUSIONS

A mode space method for TB NEGF simulations of armchair GNR FETs including phonon scattering has been presented and tested with reference to both conventional and tunnel FET structures. When no disorder is included in the simulation, an efficient decoupling of the modes in different groups (UGMS) can be employed with

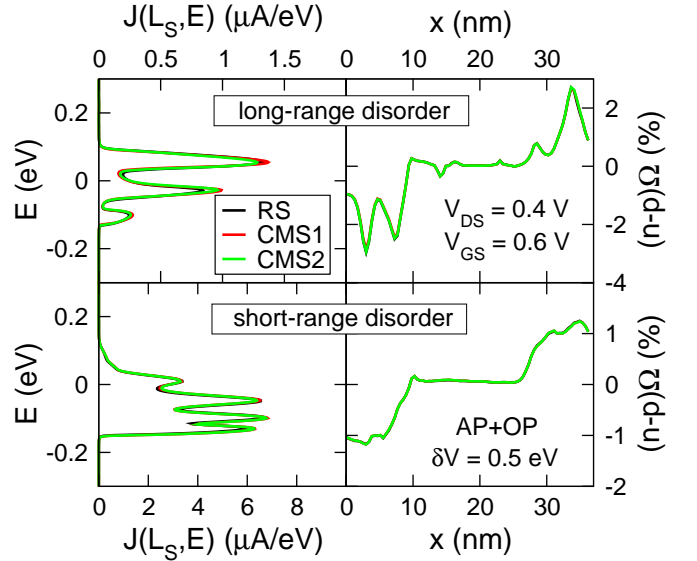


FIG. 12. Top: current spectrum at the source-channel junction (left) and average over the slab of the net electron concentration per carbon atom (right) for the device in Fig. 6–top at $V_{GS} = 0.6$ V and $V_{DS} = 0.4$ V, in the presence of a long-range disorder potential and with both AP and OP scattering. The solution method for each curve is indicated in the legend. Bottom: same as in (top) but in the presence of a short-range disorder potential.

excellent accuracy. Despite the decoupling, the method correctly accounts for inter-subband scattering. Simplified expressions of the scattering-self energies have been compared. The one obtained by neglecting some entries of the form-factor is found to be accurate except for the bias points where transport occurs by phonon-assisted BTBT. On the other hand, the real part of the scattering self-energy has only a limited effect on the device characteristics, especially for the case of optical phonon scattering, where its calculation is most demanding. In the presence of a disorder potential, the modes need to be coupled in a single group (CMS) to account for mode mixing, but no additional modes, compared to the ones used to simulate the case without disorder, need to be included to achieve accurate results.

While the computational advantage of CMS over RS is about a constant factor with respect to the GNR width (equal to about 30 for the simulation parameters chosen in this paper), the speed-up of UGMS compared to RS is about 40 for a device width of 1.5 nm and increases proportionally to the second power of the GNR width (for the 5-nm-wide device considered in this paper such speed-up is about 360).

ACKNOWLEDGMENTS

R. G. would like to thank Dr. E. Baravelli of University of Bologna for fruitful discussions on the mode-space ap-

proach. This work has been supported by the EU project GRADE 317839. The authors acknowledge the CINECA Award N. HP10CPFJ69, 2011 for the availability of high performance computing resources and support.

Appendix A: Calculation of the retarded phonon self-energy

We consider here only the RS case, the generalization of the expressions to MS being straightforward.

Replacing (4) and (5) into the last member of (7) and using the identity analogous to (7) valid for the \mathbf{G} matrices, one can write

$$\Sigma_P^R(E) = \Sigma_{AP}^R(E) + \Sigma_{OP}^R(E) \quad (\text{A1})$$

with

$$\Sigma_{AP}^R(E) = D_{ap} \mathbf{I} \circ \mathbf{G}^R(E) \quad (\text{A2})$$

and

$$\begin{aligned} \Sigma_{OP}^R(E) = D_{op} \mathbf{I} \circ \left\{ \right. \\ (N_{op} + 1) \mathbf{G}^R(E - \hbar\omega_{op}) + N_{op} \mathbf{G}^R(E + \hbar\omega_{op}) + \\ \left. + \frac{\mathbf{G}^<(E - \hbar\omega_{op}) - \mathbf{G}^<(E + \hbar\omega_{op})}{2} + \right. \\ \left. + i\text{P} \int_{-\infty}^{+\infty} \frac{\mathbf{G}^<(E' - \hbar\omega_{op}) - \mathbf{G}^<(E' + \hbar\omega_{op})}{2\pi(E - E')} dE' \right\} \quad (\text{A3}) \end{aligned}$$

which does not contain $\mathbf{G}^>$. It is worth noticing that the principal part integral in (A3) contains only a fraction of the Hermitian part of Σ_{OP}^R [19]. It is calculated here by means of a piecewise constant approximation of $\mathbf{G}^<(E')$ over the energy domain $[E_{\min}, E_{\max}]$, which is discretized uniformly with energy steps Δ_E typically of the order of $\hbar\omega_{op}/100$, i.e.

$$\begin{aligned} \text{P} \int_{-\infty}^{+\infty} \frac{\mathbf{G}^<(E' \mp \hbar\omega_{op})}{E - E'} dE' \simeq \\ \simeq \sum_j \mathbf{G}^<(E_j) \ln \left| \frac{E - E_j + \Delta_E/2 \mp \hbar\omega_{op}}{E - E_j - \Delta_E/2 \mp \hbar\omega_{op}} \right| \quad (\text{A4}) \end{aligned}$$

where the summation extends over all discrete energy points $E_j \in [E_{\min}, E_{\max}]$. An expression of Σ_{OP}^R analogous to (A3) containing $\mathbf{G}^>$ instead of $\mathbf{G}^<$ can also be derived. Expression (A3) is preferred for E higher than the contact Fermi energies, since the numerical error introduced by truncating the upper limit of the integral to E_{\max} is minimized due to the decaying nature of $\mathbf{G}^<$ at high energies. On the contrary, the alternative expression of Σ_{OP}^R which depends on $\mathbf{G}^>$ is used for low energies below the contact Fermi levels, for analogous reasons. For intermediate energies an average of the two formulations is used.

In Sec. III A, the results obtained with the full expression of Σ_{AP}^R and Σ_{OP}^R are compared with approximate solutions obtained by neglecting the respective Hermitian parts, i.e.

$$\Sigma_{AP}^R(E) \simeq D_{ap} \mathbf{I} \circ \frac{\mathbf{G}^>(E) - \mathbf{G}^<(E)}{2} \quad (\text{A5})$$

and

$$\begin{aligned} \Sigma_{OP}^R(E) \simeq \frac{D_{op}}{2} \mathbf{I} \circ \left\{ (N_{op} + 1) \mathbf{G}^>(E - \hbar\omega_{op}) + \right. \\ \left. + N_{op} \mathbf{G}^>(E + \hbar\omega_{op}) - (N_{op} + 1) \mathbf{G}^<(E + \hbar\omega_{op}) + \right. \\ \left. - N_{op} \mathbf{G}^<(E - \hbar\omega_{op}) \right\} \quad (\text{A6}) \end{aligned}$$

Expression (A6) is much more efficient than (A3) due to the absence of the integral term.

-
- [1] M. Y. Han, B. Özyilmaz, Y. Zhang, and P. Kim, Phys. Rev. Lett. **98**, 206805 (2007).
 - [2] Z. Chen, Y.-M. Lin, M. J. Rooks, and P. Avouris, Physica E **40**, 228 (2007).
 - [3] X. Wang, Y. Ouyang, X. Li, H. Wang, J. Guo, and H. Dai, Phys. Rev. Lett. **100**, 206803 (2008).
 - [4] G. Fiori and G. Iannaccone, IEEE Electron Device Letters **28**, 760 (2007).
 - [5] Y. Ouyang, Y. Yoon, and J. Guo, IEEE Trans. Electron Devices **54**, 2223 (2007).
 - [6] G. Liang, N. Neophytou, M. S. Lundstrom, and D. E. Nikonov, J. Appl. Phys. **102**, 054307 (2007).
 - [7] Y. Yoon, G. Fiori, S. Hong, G. Iannaccone, and J. Guo, IEEE Trans. Electron Devices **55**, 2314 (2008).
 - [8] P. Zhao, J. Chauhan, and J. Guo, Nano Lett. **9**, 684 (2009).
 - [9] R. Grassi, A. Gnudi, E. Gnani, S. Reggiani, and G. Bacarani, J. Comput. Electronics **8**, 441 (2009).
 - [10] S. Datta, *Quantum Transport: Atom to Transistor* (Cambridge University Press, Cambridge, UK, 2005).

- [11] Y. Ouyang, X. Wang, H. Dai, and J. Guo, Appl. Phys. Lett. **92**, 243124 (2008).
- [12] Y. Yoon, D. E. Nikonov, and S. Salahuddin, Appl. Phys. Lett. **98**, 203503 (2011).
- [13] Y. Yoon and S. Salahuddin, Appl. Phys. Lett. **101**, 263501 (2012).
- [14] N. D. Akhavan, G. Jolley, G. A. Umana-Membreno, J. Antoszewski, and L. Faraone, J. Appl. Phys. **112**, 094505 (2012).
- [15] J. Wang, E. Polizzi, and M. Lundstrom, J. Appl. Phys. **96**, 2192 (2004).
- [16] M. Luisier, A. Schenk, and W. Fichtner, J. Appl. Phys. **100**, 043713 (2006).
- [17] S. Jin, Y. J. Park, and H. S. Min, J. Appl. Phys. **99**, 123719 (2006).
- [18] S. Poli, *Modelling and simulations of postCMOS devices*, Ph.D. thesis, University of Bologna (2009).
- [19] A. Esposito, M. Frey, and A. Schenk, J. Comput. Electron. **8**, 336 (2009).
- [20] D. Nikonov, H. Pal, and G. Bouri-anoff, “Scattering in NEGF: Made simple,” <http://nanohub.org/resources/7772> (2009).
- [21] A. Afzalian, J. Appl. Phys. **110**, 094517 (2011).
- [22] M. Shin, J. Appl. Phys. **106**, 054505 (2009).
- [23] R. Grassi, A. Gnudi, E. Gnani, S. Reggiani, and G. Bacarani, IEEE Transactions on Nanotechnology **10**, 371 (2011).
- [24] P. Zhao and J. Guo, J. Appl. Phys. **105**, 034503 (2009).
- [25] Y.-W. Son, M. L. Cohen, and S. G. Louie, Phys. Rev. Lett. **97**, 216803 (2006).
- [26] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, J. Appl. Phys. **81**, 7845 (1997).
- [27] A. Svizhenko, M. P. Anantram, T. R. Govindan, B. Biegel, and R. Venugopal, J. Appl. Phys. **91**, 2343 (2002).
- [28] S. O. Koswatta, M. S. Lundstrom, M. P. Anantram, and D. E. Nikonov, Appl. Phys. Lett. **87**, 253107 (2005).
- [29] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge, UK, 1997).
- [30] A. Svizhenko and M. P. Anantram, Phys. Rev. B **72**, 085430 (2005).
- [31] M. Poljak, E. B. Song, M. Wang, T. Suligoj, and K. L. Wang, IEEE Trans. Electron Devices **59**, 3231 (2012).
- [32] A. Lherbier, B. Biel, Y.-M. Niquet, and S. Roche, Phys. Rev. Lett. **100**, 036803 (2008).